

Perceptual analysis of the speech intelligibility and soundscape of multilingual environments

Citation for published version:

Kitapci, K & Galbrun, L 2019, 'Perceptual analysis of the speech intelligibility and soundscape of multilingual environments', *Applied Acoustics*, vol. 151, pp. 124-136. <https://doi.org/10.1016/j.apacoust.2019.03.001>

Digital Object Identifier (DOI):

[10.1016/j.apacoust.2019.03.001](https://doi.org/10.1016/j.apacoust.2019.03.001)

Link:

[Link to publication record in Heriot-Watt Research Portal](#)

Document Version:

Peer reviewed version

Published In:

Applied Acoustics

General rights

Copyright for the publications made accessible via Heriot-Watt Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

Heriot-Watt University has made every reasonable effort to ensure that the content in Heriot-Watt Research Portal complies with UK legislation. If you believe that the public display of this file breaches copyright please contact open.access@hw.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Perceptual analysis of the speech intelligibility and soundscape of multilingual environments

Kivanc Kitapci^{a,*} and Laurent Galbrun^b

^a Department of Interior Architecture, Cankaya University, Ankara, Turkey

^b School of Energy, Geoscience, Infrastructure, and Society, Heriot-Watt University, Edinburgh, UK

Abstract

This paper examines the perceived speech intelligibility of English, Polish, Arabic, and Mandarin and, more generally, the soundscape associated to multilingual environments. Listening tests were used to evaluate three acoustic environments (an airport, a hospital, and a café) under three room acoustic conditions defined by a different speech transmission index (STI) (STI=0.4, 0.5 and 0.6). In the tests, participants rated eleven semantic attributes representative of speech perception and the overall soundscape (speech intelligibility, speech level, speech pleasantness, noisiness, annoyance, relaxation, comfort, environment pleasantness, eventfulness, excitement, and familiarity). Results obtained indicate that inter-language comparisons based on perceived speech intelligibility are different from those obtained from objective speech intelligibility tests. Noticeably, English participants were found to be most sensitive to changes in room acoustic conditions and to meaningful and distractive noise sources, whilst Arab participants were least sensitive to changes in room acoustic conditions and more tolerant to noise. Perceived speech intelligibility correlated significantly with non-acoustical factors (speech pleasantness, comfort and environment pleasantness), and 'emotional factors' (annoyance, relaxation, comfort and environment pleasantness) explained a large portion of the variance in soundscape assessment. Results also showed that language affected the perceived speech intelligibility marginally ($p = 0.051$) and noisiness significantly ($p = 0.047$), the latter being the best indicator of cultural variations amongst the attributes tested. Overall, the study shows that designing for speech intelligibility cannot be solely based on room acoustic parameters, especially in the case of multi-lingual environments.

1. Introduction

The effects of globalisation, population flows between nations, technology, and the new political and economic landscape of different parts of the world caused significant linguistic, cultural and demographic changes. This phenomenon increased the international interest in multilingualism [1]. The interaction between multilingual and multicultural people in public, commercial and social spaces is gaining importance, and communication is at the centre of this interaction. The aim of this study is to find out possible relations between speech intelligibility and multi-lingual communication, in terms of acoustics, linguistics, and perceptual factors related to the soundscape. To investigate the multi-dimensional structure of the intelligibility of speech in multi-lingual spaces, the project carried out was divided into two main phases.

The first phase of the project investigated the interaction of room acoustic conditions with the speech intelligibility of English, Polish, Arabic and Mandarin [2]. The results obtained indicated that there was a statistically significant difference between the word intelligibility scores of languages under all room acoustic conditions, apart from the excellent room acoustic condition corresponding to a very high speech transmission index (STI = 0.8). English was the most intelligible language under all conditions, and differences with other languages were larger when acoustic conditions were poor. Sentence intelligibility scores confirmed variations between languages, but these variations were statistically significant only at the STI = 0.4 condition (sentence tests being less sensitive to very good and very poor room acoustic conditions). Overall, the results showed that large variations between the speech intelligibility of different languages can occur, especially for spaces that are expected to be challenging in terms of room acoustic conditions.

The current paper presents results obtained from the second phase of the study, which investigated the perceived intelligibility of English, Polish, Arabic, and Mandarin, as well as the soundscape assessment of the environments examined, in order to identify potential correlations between soundscape assessment and perceived speech intelligibility. Tests were carried out for three multi-lingual spaces (an airport, a hospital, and a café) tested under three room acoustic conditions defined by a different speech transmission index (STI = 0.4, STI = 0.5, and STI = 0.6). The results obtained from both phases provide an insight into the main factors (objective and subjective) affecting multilingual communication.

* Corresponding author at: Department of Interior Architecture, Faculty of Architecture, Cankaya University, Ogretmenler Caddesi, No.14, Yuzuncuyil, Ankara 06530, Turkey
E-mail address: kivanckitapci@cankaya.edu.tr

Comparisons of objective speech intelligibility and perceived speech intelligibility are very limited in the literature. Studies often refer to “subjective speech intelligibility” when talking about tests based on word and/or sentence lists, but it is important to note that these are effectively objective tests that are not comparable to the perceived speech intelligibility examined here. The most relevant comparison between objective and perceived speech intelligibility was made by Cox et al. [3], where the perceived intelligibility corresponded to an estimated percentage of words understood. Results showed a very good correlation between objective and subjectively estimated speech intelligibility, for participants with normal hearing (correlation coefficient $r = 0.82$) [3].

Socio-lingual and cultural factors can affect communication in multilingual spaces. Although there are many recent studies on socio-linguistics and multilingual communication, most of these are not relevant to the research presented. These mainly highlight the significance of multi-lingual environments in the West (see for example [4] and [5]), but do not touch on acoustical factors. In order to fill this gap, the perception of the sound environment can be considered in addition to room acoustic conditions. The combination of physical and perceptual factors can be taken into account by the soundscape approach developed by Schafer [6], which considers all the sounds present within a space and the perception of that sound environment. The soundscape methodology is therefore a valuable approach which was used in the present study, in order to evaluate the multiple factors affecting multi-lingual communication.

Particularly relevant to the present study is how the soundscape might be rated differently across different cultures, as a large portion of the results presented here relate to soundscape assessment, rather than speech intelligibility only. In that respect, Yang and Kang [7] found that noisiness of urban open public spaces can be rated differently in different countries. In particular, results indicated that people from a noisy home environment rate noisiness lower, as they might adapt more to noisy spaces. Cultural and life style differences might also play a role in that respect, although it should be noted that a thorough review of the qualitative aspects of sound affecting annoyance points at contradictory findings in relation to cultural effects [8]. Recently, Jeon et al. [9] identified socio-cultural differences in soundscape assessments, in particular in the rating of eventfulness and human sounds, a finding that is particularly relevant to the current study, as the work was looking at a variety of soundscape attributes rather than annoyance only.

The review of previous work shows that speech intelligibility is typically assessed objectively and is mostly examined in relation to room acoustic conditions, whilst multilingualism tends to be studied from a sociological perspective. The present study aimed to expand the understanding of multilingual communication by examining perceived speech intelligibility and multilingualism with the help of the soundscape theory. More specifically, the main objectives of the research were:

- 1) To compare objective speech intelligibility [2] with perceived speech intelligibility for languages representative of a wide range of linguistic properties (English, Polish, Arabic and Mandarin).
- 2) To examine the impact of soundscape assessment on perceived speech intelligibility, and more generally on multilingual environments.
- 3) To identify the significance of differences in assessment due to language.

2. METHODOLOGY

This section describes the selection of languages, the acoustic environments, and the listening tests used in the research. All the listening tests were conducted in three multi-lingual spaces (an airport, a hospital, and a café), under three room acoustic conditions that were defined in terms of different speech transmission index values (STI = 0.4, 0.5, and 0.6).

2.1 Selecting the languages

English, Mandarin, Arabic, and Polish were selected based on five criteria: being representative of a multi-lingual environment in a western city, consonant-to-vowel ratio, tonal properties, native speakers’ population, and availability of subjects.

As the current research is applied rather than theoretical, being representative of a multilingual environment was one of the main concerns. Secondly, a significant variability between the consonant-to-vowel ratios of the languages was aimed for, as the speech intelligibility is affected by the loss of consonants [10], and as such variability would allow examining whether languages with a high consonant-to-vowel ratio are more sensitive to poor room acoustic conditions. The consonant-to-vowel ratio of languages is calculated from consonant and vowel inventories which are elements of phonology of a language. The total numbers of such sounds create the consonant and vowel inventories. Depending on the language, the number of consonants in a consonant inventory varies between 6 and 122, and the number of vowels in a vowel inventory varies between 2 and 14 [11]. Consonant-to-vowel ratios are calculated by dividing the number of consonants by the number of vowels in an inventory, resulting in a number between 1 and 29. The consonant-to-vowel ratios are divided into 5 categories, which have been used when selecting the languages of

the research presented: low (smaller than or equal to 2), moderately low (between 2 and 2.75), average (between 2.75 and 4.5), moderately high (between 4.5 and 6.5), and high (larger than or equal to 6.5) consonant-to-vowel ratio [11].

Tonality was identified as a linguistic factor that can clearly differentiate languages [12], which is why at least one tonal language had to be selected. Tone is the change of the meaning of a word by the change of pitch, and in that respect, languages can be subdivided into three categories: no tones, simple tonal system, and complex tonal system [12].

The native speakers' population of each language also had to be taken into account. The research should in fact be representative of a wide range of people; therefore, the languages with higher native speaker populations were selected. Lastly, the availability of native speakers for the selected languages was also considered, and the languages selected had to comply with high number of participants that could be found at Heriot-Watt University.

Based on the above-mentioned criteria of real environment depiction, consonant-to-vowel ratio, tonal properties, native speaker population, and availability of subjects, English (low consonant-to-vowel ratio, wide-spread usage around the world), Mandarin (complex toned system, high native speaker population), Arabic (moderately high consonant-to-vowel ratio, high native speaker population), and Polish (high consonant-to-vowel ratio, and availability of speakers) were selected.

2.2 The participants and acoustic environments

Fifteen participants per language (i.e. a total of sixty) were asked to subjectively evaluate three acoustic environments by answering eleven questions on a five-point semantic scale, under three room acoustic conditions (STI = 0.4, 0.5, and 0.6). All the listeners used as participants had one native language only, and most of the Polish, Arabic and Chinese participants were students who had been living in their native country until recently. However, these participants also knew English, and this might have affected their intelligibility scores at the lower STI levels tested (see Ref. [2] for further discussions about this limitation). Furthermore, according to the STI qualification ratings of ISO 9921 [13], the STIs used corresponded respectively to "bad", "fair" and "good" speech intelligibility conditions (Bad: STI 0-0.3; Poor: STI 0.3-0.45; Fair: STI 0.45-0.6; Good: STI 0.6-0.75; Excellent: STI 0.75-1.0).

The following criteria were applied for selecting the acoustic environments: oral communication must be at the centre of attention, the environments should represent a variety of acoustic conditions; and the test participants should have an experience of the selected environments.

Three multi-lingual environments were selected for the study. The first case selected was an airport check-in area, as an example of a high reverberation time and high background noise acoustic environment. Airports are common public environments in most global cities, where oral communication between a passenger and a check-in desk attendant is often crucial. The simulated airport enclosure is typically large and spacious with a high ceiling, therefore, leading to a reverberant acoustic environment. The background noise is typically fairly steady (mainly hubbub speech), and occasionally there are public announcements (PA) and other impact sounds (e.g. footsteps and luggage wheels).

The second case selected was a hospital reception area. The speech content of a hospital reception area is usually crucial since the context is about health issues. Conversations between a patient and a receptionist can accommodate critical information, which cannot be risked being unintelligible. Compared to the other public spaces selected for the experiments, the hospital reception area simulated was a medium sized enclosure leading to a medium to low reverberation time, with a relatively low background noise mostly composed of hubbub speech noise. A telephone ringing was also present.

The last case selected was a café. Although the speech content in a café environment is not as crucial as the other cases, conversation is still at the centre of attention. Additionally, especially in global cities, cafés are one of the most multi-cultural and multi-lingual public spaces. It is also a relaxed environment, as opposed to the stressful environments represented by the airport check-in area and the hospital reception. The simulated café environment considered in this study was a medium to large sized space, with a moderately-high reverberation time and continuous background noise, which was mostly composed of hubbub speech noise.

2.3 Listening tests

The first step in preparing the audio-visual materials was deciding on the room acoustic conditions that would be tested. After analysing the data from the first phase of the study [2], the largest variation of word/sentence intelligibility scores was observed at STI = 0.4. In practice, this represents a poor room acoustic condition [13] and most spaces should be expected to perform with higher STI values. Therefore, three STI values were aimed, which were STI = 0.4, 0.5 and 0.6. This guaranteed investigating conditions where variations in intelligibility between languages are highest, but representative of real cases (i.e. lower STI conditions are rare and STI conditions above 0.6 are not expected to show significant differences between languages).

The second step was designing the sentence lists. Different sentence lists were prepared for each of the three cases. Each case contained six sentences that were uniquely designed to match the context of the environment. For English, each sentence was made of approximately 50 syllables (minimum 44 syllables and maximum of 51 syllables). The sentences were then translated to Polish, Arabic, and Mandarin by native speakers of the languages. During the translation, attention was given to the syllable counts of the sentences to be comparable within the languages. It was also important that the listeners felt as part of the conversation; therefore, the sentences were designed to simulate active engagement of the participants, either by directing a question, or by illustrating a task. Examples of the sentence lists used are presented in Table 1.

Table 1 Example list of the English sentences.

Airport check-in area	I am afraid the luggage allowed on this flight is two pieces maximum, regardless of the maximum weight permitted. The charge per extra luggage is fifteen Euros, which you can pay at the airline's counter.
Hospital reception area	In order to book an appointment, I first need you to fill in this form and submit it to me when completed. Please write down your name, date of birth, phone number and health insurance number if available.
Café	I am really looking forward to the weekend. Yesterday I spent some time planning a two-hour hike in the mountains, as well as a short boat trip on the lake, if the weather is good. Would you be interested in coming with me?

The third step was recording the sentences and identifying the background noise samples to be used in the final audio files. The lists were recorded in the anechoic chamber of Heriot-Watt University, using four native speakers of each language (2 males and 2 females). Because of the significant variety of accents within languages, attention was given to the origin of the speakers. The English speakers had to speak English with Received Pronunciation (RP) [14], which is normally associated with formal speech and tends to be spoken in the south of England. The Arabic speakers were selected from Syria, although the origin of Arabic speakers was not crucial, as the Arabic material was written and recorded in modern standard Arabic (al-fuṣḥá), for which the pronunciation is independent from accents and dialects [15]. Care was also taken in the selection of Polish and Mandarin speakers, so that they could produce formal speech material.

The airport and hospital background noise samples could not be recorded at actual locations because of security restrictions; therefore, previously recorded high-quality sound samples were used. After subjectively reviewing the catalogue of the 'audiosparx.com' website in terms of audio quality, sample length, and the availability of sound marks related to the environments, one background noise sample was selected for each of the environments. The airport and hospital background noises were 24 seconds long samples that were selected out of longer audio recordings of an airport and a hospital. Both of the audio files were high-quality wave sound files (44.1 kHz, 16 bit). The café background noise sample was recorded at a canteen of Heriot-Watt University, a medium-large sized enclosure that attracts many people from the university. The café background noise sample was recorded using a digital sound recorder Zoom H4n during the lunchtime, which is the most crowded time period of the day. After reviewing the recording in terms of homogeneity, a 24 seconds long sample was selected.

The next step was mixing the speech and background noise sound samples and finalising the sound files by adding reverberation to the speech samples in order to achieve the aimed STI values for each of the three environments. Digital audio processing was carried out by using Studio One 2 audio production software (PreSonus audio electronics), installed on a personal computer (PC) connected to an external M-Audio USB sound card. Sound pressure level measurements of the speech and the background noise samples were carried out by connecting a sound level meter Brüel and Kjaer Type 2250 to the master sound output of the M-Audio USB sound card.

The STI values were computed individually for each of the 288 speech recordings (6 sentences, 3 environments, 4 speakers, and 4 languages) by using the modulation transfer function (MTF) method. The MTF method is a measure used to examine the effects of the enclosures' acoustic properties on the intelligibility of speech [16]. The speech signal is affected by two factors until reaching the listener: the signal-to-noise ratio (S/N), which is independent from the modulation frequency; and the reverberation time (T), which has a different effect on every frequency. The modulation reduction factor $m(f_m)$ defines the decrease in modulation caused by acoustical conditions and is a function of the modulation frequency [17]. The detailed procedure for calculating the STI from $m(f_m)$ can be found in Long [17].

Since the airport, hospital, and café environments varied in terms of overall volume of the spaces, the direct field contribution was included in the calculation of the modulation reduction factor m . In order to calculate the direct field contribution, the critical distance and source-to-receiver distance had to be identified for each of the acoustic environments. The critical distance is the point where the direct sound pressure level is equal to the reverberant field

sound pressure level [17], and can be computed by using the following equation, when assuming spherical propagation of sound:

$$r_c = \sqrt{\frac{A}{16\pi(1 - \bar{\alpha})}} \quad (1)$$

where A is the total absorption in the room (m^2) and $\bar{\alpha}$ is the average absorption coefficient. The source-to-receiver distance was assumed as 1 m for all the acoustic environments. After finding the critical distance and knowing the source-to-receiver distance, the modulation reduction factor m was computed by using the following equation [17]

$$m(\omega_m) = \frac{(A^2 + B^2)^{1/2}}{C} [1 + 10^{-0.1L_{SN}}]^{-1} \quad (2)$$

with

$$A = \frac{Q}{r^2} + \frac{Q}{r_c^2} \left[1 + \left(\frac{\omega_m T_{60}}{13.8} \right)^2 \right]^{-1} \quad (3)$$

$$B = \frac{\omega_m T_{60}}{13.8} \frac{Q}{r_c^2} \left[1 + \left(\frac{\omega_m T_{60}}{13.8} \right)^2 \right]^{-1} \quad (4)$$

$$C = \frac{Q}{r^2} + \frac{Q}{r_c^2} \quad (5)$$

where Q is the source directivity, r is the source-to-receiver distance (m), r_c is the critical distance (m), ω_m is the modulation angular frequency (radians/s) ($\omega_m = 2\pi f_m$), L_{SN} is the signal-to-noise ratio (dB), and T_{60} is the reverberation time (s).

As the STI values were previously decided based on the first phase results of the study, the reverberation time on the speech recordings and the signal-to-noise ratios were adjusted to achieve the desired STI values. Based on the comparative volumes of the three environments and in order to achieve a variety of reverberation times, the airport check-in area, which is the enclosure with the highest volume was modelled to have a reverberation time of 1.5 s, the café, which is the medium-large sized enclosure was modelled to have a reverberation time of 1.2 s, and the hospital reception area, which is the medium-sized enclosure was modelled to have a reverberation time of 1.0 s, across all frequencies. The reverberation time was varied using the 'Room Reverb' plugin of the Studio One 2 software (Feedback Delay Network (FDN) algorithm). The signal-to-noise ratios were then set manually by adjusting the sound pressure level of the background noise samples in order to achieve the desired STI values. The audio files were finalised by adjusting the sound pressure level of the speech sample to 65 dB(A) and mixing it with the background noise sample in order to achieve the desired signal-to-noise ratios.

Following the recording and post-processing procedure, a total of nine sound environments were created. The acoustical properties of these environments are presented in Table 2.

Table 2 Acoustical properties of the environments tested.

Environment	Reverberation Time	S/N	STI
Airport	1.5s	-2.90 dB	0.4
Airport	1.5s	0.06 dB	0.5
Airport	1.5s	3.12 dB	0.6
Hospital	1.0s	-2.33 dB	0.4
Hospital	1.0s	1.00 dB	0.5
Hospital	1.0s	5.26 dB	0.6
Café	1.2s	-2.72 dB	0.4
Café	1.2s	0.15 dB	0.5
Café	1.2s	3.11 dB	0.6

Case specific visuals were presented to the test participants together with the sound samples. For the airport check-in area and the hospital reception area, a high-resolution photograph was selected from online searches. For the café, a high-resolution photograph was shot at the canteen of Heriot-Watt University during lunch-time (same time period of the background noise recordings). Attention was given to have photographs with a general view and with no distracting focal points (Figure 1).

The audio samples and the visuals of the environments were compiled in the form of a slide-show for the listening tests. The slide shows were prepared by using the software Microsoft PowerPoint 2013. Each slide-show consisted of a total 28 slides, including detailed instructions on the listening test sessions. Both the speech samples used for each acoustic environment and the order of the acoustic environments presented to each participant were randomised to avoid order effects.

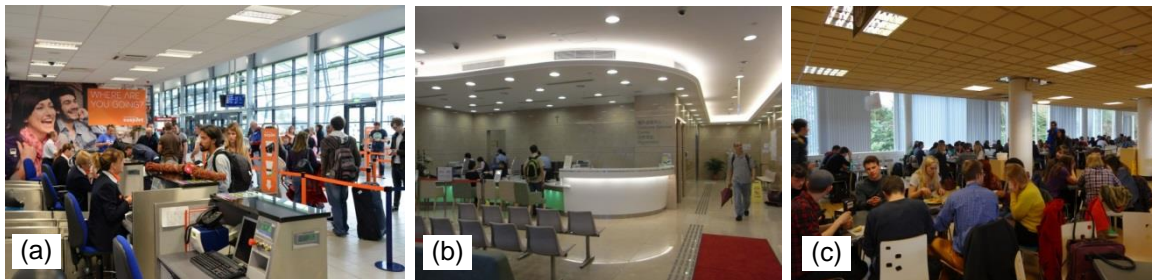


Figure 1 The visuals of the environments tested. (a) Airport check-in area: London Southend Airport, UK (Source: Google Images (n.d.)). (b) Hospital reception area: Saint Paul's hospital lobby, Hong-Kong (Source: Google Images (n.d.)). (c) Café: Heriot-Watt University, Edinburgh, UK.

The subjects were asked to evaluate the audio-visual material through semantic differential analysis. The semantic differential technique was first developed by Osgood et al. [18] to identify emotional meanings of the words. In the previous literature semantic differential analysis has been adopted for soundscape analysis by identifying sounds and their linguistic and psychological meanings [19]. Detailed instructions were presented both on-screen and orally prior to the listening tests. Participants were allowed to listen to each acoustic environment only once. Each evaluation form consisted of 11 5-point scale semantic questions (3 semantic questions for assessing the speech and 8 semantic questions for assessing the overall acoustic environment). In line with the review of previous work carried out in ref. [20], the attributes tested were speech intelligibility, speech loudness, speech pleasantness, noisiness, annoyance, relaxation, acoustic comfort, environment pleasantness, eventfulness, excitement, and familiarity. All the semantic attributes presented were based on a -2 to +2 range (e.g.: -2 = very unintelligible; -1 = unintelligible; 0 = neither intelligible nor unintelligible; +1 = intelligible; +2 = very intelligible), and no endpoints were used in the scales. After completing the evaluation form, participants were asked to proceed to the next acoustic environment, and the process was repeated until all of the nine cases listed in Table 2 had been tested.

3. RESULTS

In this section, results of the semantic differential analysis and principal component analysis (PCA) are presented and analysed. A total of four statistical analysis methods were applied to the data sets in order to test several hypotheses; these were Intra-Class Correlation analysis (ICC), one-way Analysis of Variance (one-way ANOVA), repeated measures Analysis of Variance (repeated measures ANOVA), and Principal Component Analysis (PCA). All of these were computed using the Statistical Package for Social Sciences (SPSS) v.22 software.

Consistency between the participants taking part in the tests was analysed by using the Intra-Class Correlation analysis [21]. In order to assess between subject's reliability, the Intra-Class Correlation Coefficient (ICC) was computed for the participants of each language. The absolute agreement average measures ICC analysis with the two-way mixed model revealed that the answers of participants were in agreement for English (ICC = 0.924), Mandarin (ICC = 0.898), Arabic (ICC = 0.912), and Polish (ICC = 0.881), where $ICC > 0.720$ is usually considered as an acceptable value for social sciences [22]. This confirms that the use of 15 listeners per language was appropriate and that the results presented are reliable.

3.1 Semantic differential analysis

In this section, results of the semantic differential analysis are presented. As the key attribute tested, perceived speech intelligibility is analysed first, followed by the analysis of the other ten semantic attributes.

3.1.1 Perceived speech intelligibility

Fifteen participants per language (i.e. a total of sixty) were asked to subjectively evaluate intelligibility by answering a five-point semantic scale, from very unintelligible (-2) to very intelligible (+2), under three room acoustic conditions, in three digitally simulated multi-lingual environments (i.e., nine cases were rated by each participant).

Figure 2 and Table 3 show the relationship between the intelligibility attribute scores and the STI levels at the airport, the hospital, and the café respectively. Logarithmic regression lines are displayed in Figure 2, as those provided the best fits (highest R^2 values) for most cases. Logarithmic regressions have been used even for the few cases where linear regressions provided better fits, simply because the latter would be misleading when extrapolating data at high and low STIs: previous work has demonstrated that intelligibility goes towards a plateau at high STI values, whilst it decreases rapidly towards low STI values, i.e. its behaviour is logarithmic (see Annex F of ref. [13]).

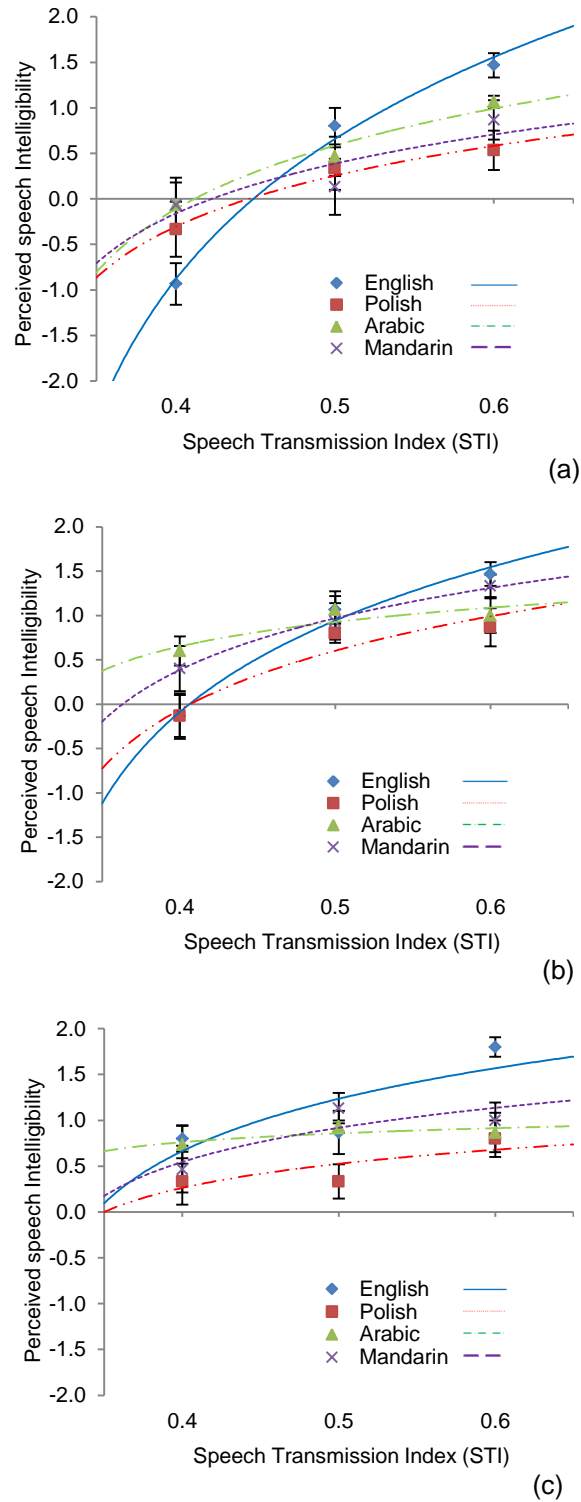


Figure 2 Perceived speech intelligibility scores of English, Polish, Arabic and Mandarin for the three environments tested (data markers, standard errors of the mean and logarithmic regression lines). (a) Airport. (b) Hospital. (c) Café.

Table 3 Perceived speech intelligibility scores at the airport, hospital, and café.

Environment	STI	English	Polish	Arabic	Mandarin
Airport	0.4	-0.93	-0.33	-0.07	-0.07
	0.5	0.80	0.33	0.47	0.13
	0.6	1.47	0.53	1.07	0.87
Hospital	0.4	-0.13	-0.13	0.60	0.40
	0.5	1.07	0.80	1.07	0.93
	0.6	1.47	0.87	1.00	1.33
Café	0.4	0.80	0.33	0.73	0.47
	0.5	0.87	0.33	0.93	1.13
	0.6	1.80	0.80	0.87	1.00

Results show that the scores vary between the STI conditions, between the environments and between the languages. As expected, the perceived intelligibility scores tend to increase as the STI increases, i.e. as room acoustic conditions improve. Regarding language scores, the most noticeable results are that English was perceived to be the most intelligible language across all conditions, apart from the airport and the hospital at STI = 0.4 (i.e. most intelligible in 7 out of 9 conditions), whilst Polish was consistently rated as the least intelligible language (with the exception of the airport at STI = 0.4 and STI = 0.5). Regarding environments, it can be noted that intelligibility tended to be consistently better in the café at STI = 0.4.

When these results are compared with the results of the first phase of the study [2], contradictions are observed. For instance, the sentence scores of English from the first phase of the study showed that English was the second most intelligible language at STI = 0.4. The word intelligibility scores also revealed that it was the most intelligible language under all the acoustic conditions tested. However, in the results presented here English was perceived to be the least intelligible language at STI = 0.4 in 2 out of 3 cases (airport and hospital, which include distractive noise sources), but the most intelligible language at the café. In other words, similar results of objective and perceived speech intelligibility were only observed at the café.

When Polish results are compared to the first phase of the study [2], contradictions are also observed. The sentence intelligibility scores from the first phase of the study revealed that Polish was the most intelligible language at STI=0.6; and according to the word intelligibility scores of the first phase of the study, Polish was the second most intelligible language at STI = 0.6. However, in the results presented here it was the least intelligible language at STI = 0.6 for the three environments tested.

The perceived intelligibility scores of Arabic and Mandarin were found to be similar, again contradicting objective intelligibility findings from the first phase of the study, where Arabic showed consistently lower intelligibility in both word and sentence tests, whilst mandarin tended to show better objective intelligibility [2]. The first phase word and sentence intelligibility scores of Arabic were the lowest at STI = 0.4 and STI = 0.6; however, the perceived speech intelligibility scores of Arabic were the highest in 2 out of 3 cases at STI = 0.4 (airport and hospital). The rankings were low at STI = 0.6, but not the lowest, as Arabic had the second highest intelligibility score at the airport, and the second lowest intelligibility score at the hospital and at the café. In the first phase of the study [2], Mandarin had the highest sentence intelligibility scores and the second highest word intelligibility scores at STI = 0.4, which complies with the perceived intelligibility attribute scores obtained for the airport only. Furthermore, the perceived intelligibility of Mandarin was the lowest at STI = 0.5 at the airport and at the café.

In order to quantify variability in perceived speech intelligibility with room acoustic conditions, Table 4 presents the differences between the highest and the lowest intelligibility attribute scores of each language at the airport, the hospital and the café. In this table it can be seen that English showed the largest variance for all the environments, whilst Arabic had the lowest variance in 2 out of 3 cases (hospital and café). These results indicate that English participants were most affected by changes in room acoustic conditions, whilst Arabic participants tended to be the least affected.

Figure 3 shows the perceived speech intelligibility scores (average of the 3 room acoustic conditions) for each environment and language. The average scores of Polish were always the lowest between languages. Other language results are fairly similar, with the exception of English intelligibility in the café which was found to be significantly higher.

Table 4 Differences between the lowest and the highest average intelligibility attribute scores of English, Polish, Arabic, and Mandarin at the airport, the hospital, and the café.

Environment	English	Polish	Arabic	Mandarin
Airport	2.40	0.86	1.12	0.92
Hospital	1.59	0.99	0.40	0.93
Café	1.00	0.47	0.13	0.54

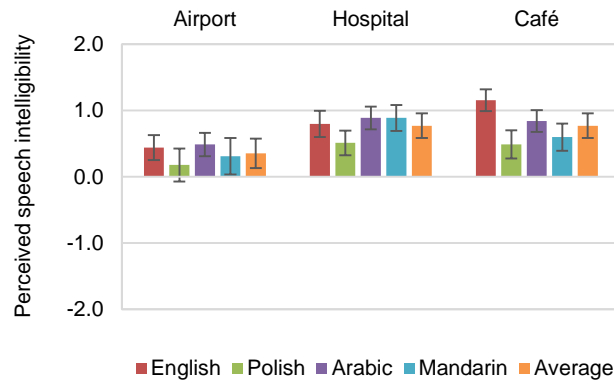


Figure 3 Perceived speech intelligibility scores (average of the 3 room acoustic conditions), for each environment and language (data bars and standard errors of the mean).

The individual and interaction effects of the three independent variables (STI, environment, and language) on perceived speech intelligibility were analysed by repeated measures ANOVA, with a confidence interval set to 95%. This allowed identifying statistically significant differences in relation to the independent variables. Sphericity of the data was tested using Mauchly's test, to check the homogeneity of variance across conditions [21]. When sphericity was not met, Greenhouse-Geisser corrections were applied. Effect sizes were computed by using eta squared (η^2).

Mauchly's test of sphericity indicated that the assumption of sphericity had been significantly violated only for STI effects ($\chi^2(2) = 7.641, p = 0.022$). The repeated measures ANOVA with a Greenhouse-Geisser correction showed that the effects of speech transmission index (STI) [$F(1.7,99.1) = 82.36, p = 0.000, \eta^2 = 0.595$] and environment [$F(1.9,108.3) = 17.25, p = 0.000, \eta^2 = 0.235$] on the intelligibility attribute were statistically significant. Additionally, marginal significance [$F(3,56) = 2.75, p = 0.051, \eta^2 = 0.129$] was observed for variations between the speech intelligibility of different languages. Furthermore, interaction effects of STI and language [$F(5.31,99.1) = 4.42, p = 0.000, \eta^2 = 0.225$], STI and environment [$F(3.41,191.1) = 6.72, p = 0.000, \eta^2 = 0.107$], and STI, environment and language [$F(10.24,191.08) = 2.30, p = 0.013, \eta^2 = 0.110$] were statistically significant.

The differences between the perceived intelligibility of the four languages were statistically analysed for each room acoustic condition by using one-way ANOVA. The homogeneity of variances was tested using Levene's test, with a confidence interval set to 95%. Levene's test showed that homogeneity was not met for 4 out of 9 cases. This was for the airport – STI = 0.6 [$F(3,56) = 6.49, p = 0.001$], the hospital – STI = 0.4 [$F(3,56) = 3.15, p = 0.032$], the café – STI = 0.4 [$F(3,56) = 4.66, p = 0.006$], and the café – STI = 0.5 [$F(3,56) = 4.19, p = 0.010$]. In this case, one-way Welch's ANOVA results were used, as these are not sensitive to unequal variances. The one-way Welch's ANOVA results revealed that 4 out of 9 conditions showed significant differences ($p < 0.05$) between languages. These were the airport – STI=0.6 [$F(3,27.77) = 5.04, p = 0.006, \eta^2 = 0.22$], the hospital STI=0.4 [$F(3,30.48) = 3.10, p = 0.041, \eta^2 = 0.12$], the café – STI=0.5 [$F(3,27.35) = 3.64, p = 0.025, \eta^2 = 0.16$], and the café – STI=0.6 [$F(3,29.57) = 10.62, p = 0.000, \eta^2 = 0.25$].

Key findings of perceived speech intelligibility

- There are discrepancies in inter-language comparisons based on either objective or perceived speech intelligibility.
- English participants were most sensitive to poor room acoustic conditions, and most affected by distractive noise sources.
- Arab participants were the least sensitive to room acoustic conditions.
- Perceived speech intelligibility varied mainly with the type of environment and the room acoustic conditions (quantified here by the STI).
- Language also affected perceived speech intelligibility, but the statistical significance was found to be marginal ($p = 0.051$), with 4 out of 9 cases showing statistically significant differences amongst languages ($p < 0.05$).

3.1.2 Other semantic attributes

In addition to perceived speech intelligibility, participants were also asked to evaluate 10 additional semantic attributes: loudness of speech, pleasantness of speech, noisiness, annoyance, relaxation, comfort, pleasantness of the environment, eventfulness, excitement, and familiarity. Each attribute was evaluated using a five-point semantic scale, for the three room acoustic conditions and three environments mentioned previously.

First, correlations between the attribute scores and the perceived speech intelligibility scores were analysed (Table 5). The speech loudness, speech pleasantness, comfort and environmental pleasantness were significantly correlated with the perceived intelligibility attribute scores of all the four languages tested. The noisiness, annoyance and relaxation were significantly correlated with the perceived speech intelligibility attribute scores of English, Polish, and Mandarin, but not Arabic. Eventfulness showed significant correlations with the perceived intelligibility attribute

scores of only two languages: English and Polish. Moreover, the familiarity attribute scores were correlated with the perceived intelligibility attribute scores of Mandarin only. Finally, it should be noted that the excitement attribute showed no significant correlations with the perceived speech intelligibility of any language. Further analysis of each semantic attribute is given below, where averages calculated across all STI conditions are displayed. Results obtained at each condition can be found in Ref. [23].

Table 5 Overall Spearman's correlation coefficient between the perceived speech intelligibility attribute and the other attributes across all 9 conditions (3 environments \times 3 STIs).

Attribute	English	Polish	Arabic	Mandarin
Speech Loudness	0.522**	0.244**	0.297**	0.376**
Speech pleasantness	0.325**	0.190*	0.355**	0.509**
Noisiness	-0.321**	-0.195*	-0.084	-0.300**
Annoyance	-0.551**	-0.265**	-0.065	-0.461**
Relaxation	0.487**	0.289**	0.110	0.421**
Comfort	0.457**	0.310**	0.182*	0.439**
Environment pleasantness	0.486**	0.167*	0.245**	0.487**
Eventfulness	-0.417**	-0.188*	-0.084	-0.058
Excitement	0.028	0.073	0.017	0.102
Familiarity	0.062	0.043	0.072	0.306**

** Correlation is significant at the 0.01 level

* Correlation is significant at the 0.05 level.

3.1.2.1 Speech loudness

Figure 4 shows that there are small variations in speech loudness across languages, whilst there are some variations across environments. Speech was mainly rated as not loud (negative values), and speech loudness was found to increase with the STI [23].

Mauchly's test of sphericity indicated that the assumption of sphericity had been significantly violated only for STI effects ($\chi^2(2) = 7.029$, $p = 0.030$). The repeated measures ANOVA with a Greenhouse-Geisser correction revealed that the effects of STI [$F(1.79,100.00) = 96.48$, $p = 0.000$, $\eta^2 = 0.633$] and environment [$F(1.92,107.82) = 17.08$, $p = 0.000$, $\eta^2 = 0.234$] were statistically significant on the speech loudness attribute. Additionally, interaction effects of environment and language [$F(5.78,107.82) = 2.85$, $p = 0.014$, $\eta^2 = 0.132$], and STI and environment [$F(3.76,210.54) = 5.29$, $p = 0.001$, $\eta^2 = 0.086$] were statistically significant on perceived speech loudness.

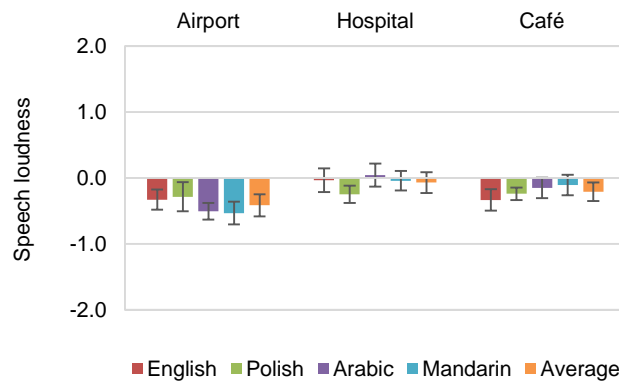


Figure 4 Speech loudness scores (average of the 3 room acoustic conditions), for each environment and language (data bars and standard errors of the mean).

3.1.2.2 Speech pleasantness

Figure 5 shows that, in most environments, speech was rated as neither pleasant nor unpleasant (close to 0 axis value), and was rated the highest at the café (least stressful environment). Speech pleasantness was found to increase with the STI [23].

Mauchly's test of sphericity indicated that the assumption of sphericity had been significantly violated only for STI effects ($\chi^2(2) = 21.916$, $p = 0.000$) and environment effects ($\chi^2(2) = 15.314$, $p = 0.000$). The repeated measures ANOVA with a Greenhouse-Geisser correction revealed that the effects of STI [$F(1.50,84.30) = 28.23$, $p = 0.000$, $\eta^2 = 0.335$] and environment [$F(1.61,90.10) = 21.12$, $p = 0.000$, $\eta^2 = 0.274$] on the speech pleasantness attribute were statistically significant.

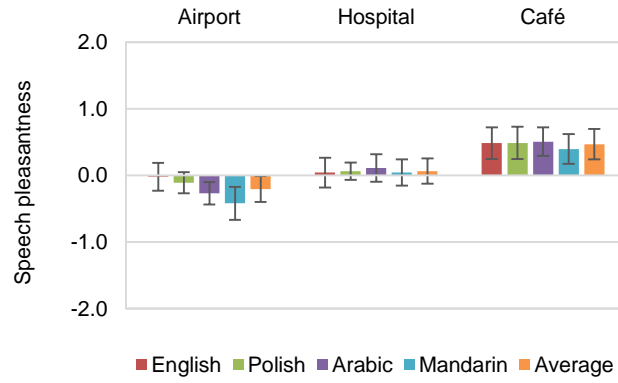


Figure 5 Speech pleasantness scores (average of the 3 room acoustic conditions), for each environment and language (data bars and standard errors of the mean).

3.1.2.3 Noisiness

Figure 6 shows that all environments were rated as noisy (positive values), and ratings were always lower for Arabic participants (i.e. less noisy). Noisiness was found to decrease with the STI [23].

Mauchly's test of sphericity was not violated. The repeated measures ANOVA with sphericity assumed revealed that the effect of languages on the noisiness attribute scores was statistically significant [$F(3,56) = 2.82, p = 0.047, \eta^2 = 0.131$], along with the effects of STI [$F(2,112) = 79.98, p = 0.000, \eta^2 = 0.588$] and the environment [$F(2,112) = 5.40, p = 0.006, \eta^2 = 0.088$]. Furthermore, the interaction effect of STI and environment was statistically significant [$F(4,224) = 3.68, p = 0.006, \eta^2 = 0.062$].

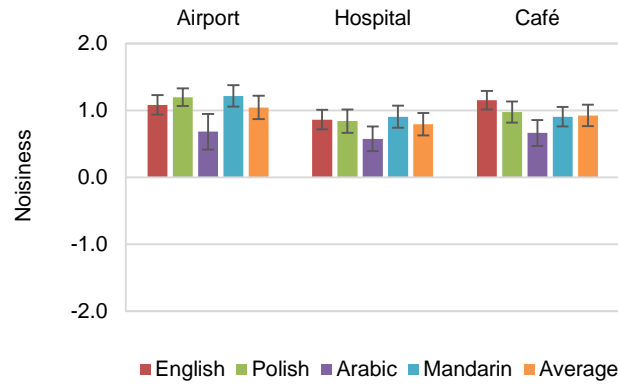


Figure 6 Noisiness scores (average of the 3 room acoustic conditions), for each environment and language (data bars and standard errors of the mean).

3.1.2.4 Annoyance

Figure 7 shows that all environments were rated as annoying (positive values), annoyance being the lowest in the café. Annoyance was found to decrease with the STI [23].

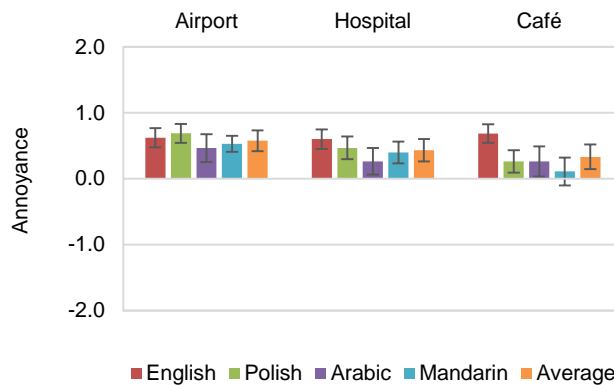


Figure 7 Annoyance scores (average of the 3 room acoustic conditions), for each environment and language (data bars and standard errors of the mean).

Mauchly's test of sphericity was not violated. The repeated measures ANOVA with sphericity assumed revealed that the effects of STI [$F(2,112) = 77.24, p = 0.000, \eta^2 = 0.058$] and environment [$F(2,112) = 6.63, p = 0.002, \eta^2 = 0.106$] on the annoyance attribute were statistically significant. Furthermore, the interaction effect of STI and environment was statistically significant [$F(4,224) = 2.20, p = 0.000, \eta^2 = 0.106$].

3.1.2.5 Relaxation

Figure 8 shows that most conditions (10 out of 12) were rated negatively (i.e. not relaxing environments). As expected, relaxation was highest in the café, and was found to increase with the STI [23].

Mauchly's test of sphericity indicated that the assumption of sphericity had been significantly violated only for environment effects ($\chi^2(2) = 8.936, p = 0.011$). The repeated measures ANOVA with a Greenhouse-Geisser correction revealed that the effects of STI [$F(1.98,110.74) = 47.70, p = 0.000, \eta^2 = 0.460$] and environment [$F(1.74,97.39) = 23.82, p = 0.000, \eta^2 = 0.298$] on the relaxation attribute were statistically significant. Additionally, the interaction effect of STI, environment and language was also statistically significant [$F(10.80,201.68) = 2.71, p = 0.003, \eta^2 = 0.127$].

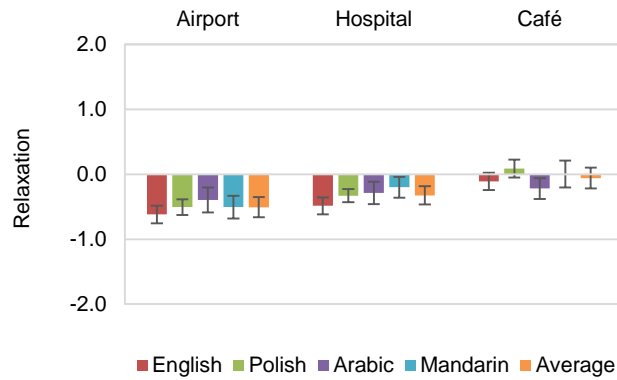


Figure 8 Relaxation scores (average of the 3 room acoustic conditions), for each environment and language (data bars and standard errors of the mean).

3.1.2.6 Comfort

Figure 9 shows that most conditions (10 out of 12) were rated negatively (i.e. not comfortable environments), comfort being the highest in the café. Comfort was found to increase with the STI [23].

Mauchly's test of sphericity was not violated. The repeated measures ANOVA with sphericity assumed revealed that the effects of STI [$F(2,112) = 62.31, p = 0.000, \eta^2 = 0.527$] and environment [$F(2,112) = 13.45, p = 0.000, \eta^2 = 0.194$] on the comfort attribute were statistically significant. It also revealed that the interaction effects of STI and language [$F(6,112) = 2.43, p = 0.030, \eta^2 = 0.115$] and environment and language [$F(6,112) = 2.91, p = 0.011, \eta^2 = 0.135$] were statistically significant.

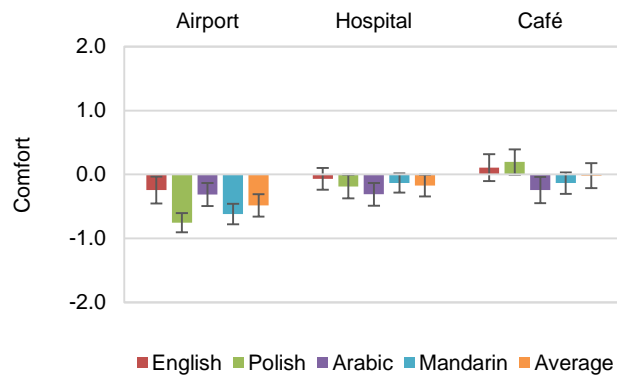


Figure 9 Comfort scores (average of the 3 room acoustic conditions), for each environment and language (data bars and standard errors of the mean).

3.1.2.7 Environment pleasantness

Figure 10 shows that almost all conditions (11 out of 12) were rated negatively (i.e. not pleasant environments), ratings being the highest in the café. The environment pleasantness was found to increase with the STI [23].

Mauchly's test of sphericity was not violated. The repeated measures ANOVA with sphericity assumed revealed that the effects of STI [$F(2,112) = 37.21, p = 0.000, \eta^2 = 0.399$] and environment [$F(2,112) = 11.13, p = 0.000, \eta^2 = 0.166$] on the environmental pleasantness attribute were statistically significant. It also revealed that the interaction effects of STI and language [$F(6,112) = 2.59, p = 0.022, \eta^2 = 0.122$], and STI and environment [$F(4,224) = 5.08, p = 0.001, \eta^2 = 0.083$] were statistically significant.

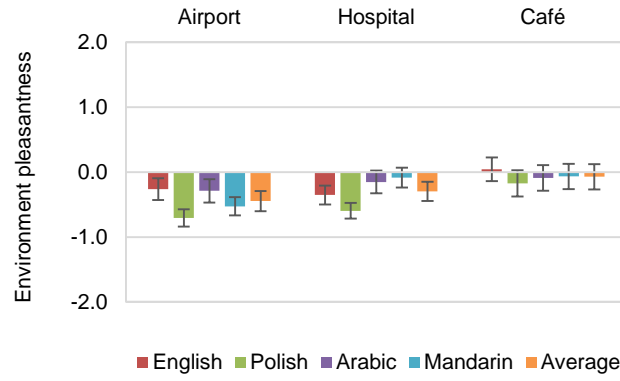


Figure 10 Environment pleasantness scores (average of the 3 room acoustic conditions), for each environment and language (data bars and standard errors of the mean).

3.1.2.8 Eventfulness

Figure 11 shows that all conditions were rated as eventful (i.e. positive values), and eventfulness was found to decrease with the STI [23].

Mauchly's test of sphericity indicated that the assumption of sphericity had been significantly violated only for environment effects ($\chi^2(2) = 10.162, p = 0.006$). The repeated measures ANOVA with a Greenhouse-Geisser correction revealed that the effects of STI [$F(1.96, 110.05) = 20.75, p = 0.000, \eta^2 = 0.270$] and environment [$F(1.71, 95.83) = 6.19, p = 0.005, \eta^2 = 0.100$] on the eventfulness attribute were statistically significant. It also revealed that the interaction effects of STI and language [$F(5.89, 110.05) = 2.74, p = 0.017, \eta^2 = 0.128$], STI and environment [$F(3.56, 199.56) = 4.27, p = 0.004, \eta^2 = 0.071$] and STI, environment and language [$F(10.69, 199.56) = 2.20, p = 0.017, \eta^2 = 0.106$] were statistically significant.

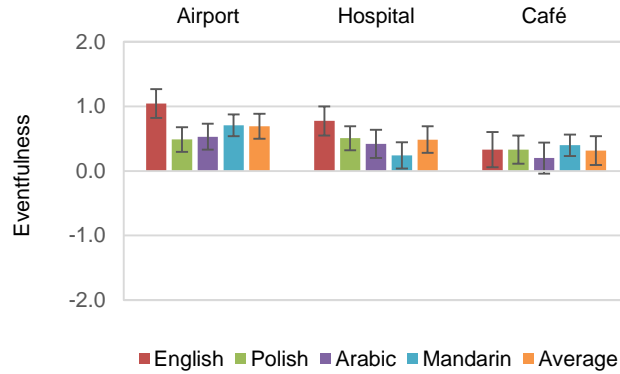


Figure 11 Eventfulness scores (average of the 3 room acoustic conditions), for each environment and language (data bars and standard errors of the mean).

3.1.2.9 Excitement

Figure 12 shows that overall conditions were rated as neither exciting nor unexciting (close to 0 axis value), and no changes with STI were observed [23].

Mauchly's test of sphericity was not violated. The repeated measures ANOVA with sphericity assumed revealed that the interaction effects of environment and language [$F(6, 112) = 2.68, p = 0.018, \eta^2 = 0.125$], and STI, environment and language [$F(12, 224) = 3.03, p = 0.001, \eta^2 = 0.140$] on the excitement attribute were statistically significant.

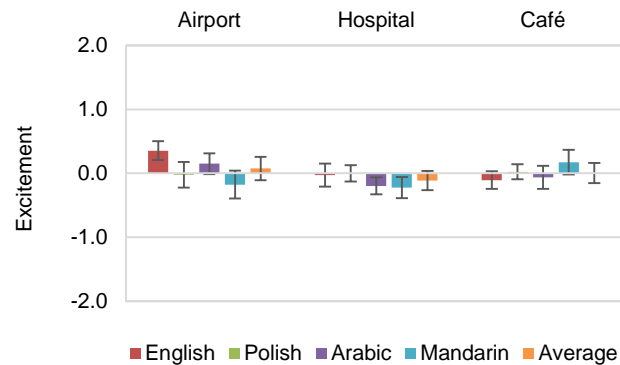


Figure 12 Excitement scores (average of the 3 room acoustic conditions), for each environment and language (data bars and standard errors of the mean).

3.1.2.10 Familiarity

Figure 13 shows that all conditions were rated as familiar (positive values), and no clear trend was found for changes in STI (varied with environment and language [23]).

Mauchly's test of sphericity indicated that the assumption of sphericity had been significantly violated only for the interaction effect of STI and environment ($\chi^2(9) = 37.459$, $p = 0.000$). The repeated measures ANOVA with a Greenhouse-Geisser correction revealed that the effect of the environment on the familiarity attribute was statistically significant [$F(1.99,111.65) = 14.76$, $p = 0.000$, $\eta^2 = 0.209$]. It also revealed that the interaction effects of STI and language [$F(5.89,109.90) = 2.89$, $p = 0.012$, $\eta^2 = 0.134$], and STI, environment and language [$F(9.00,168.02) = 2.03$, $p = 0.038$, $\eta^2 = 0.098$] were statistically significant.

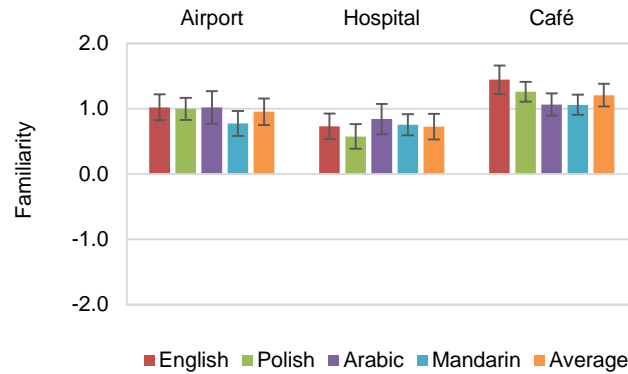


Figure 13 Familiarity scores (average of the 3 room acoustic conditions), for each environment and language (data bars and standard errors of the mean).

In order to identify which of the independent variables (STI, environment and language) had the most effect across all the semantic attributes tested, a ranking is given in Table 6. This shows that variations in results were largely due to the environment (airport vs. hospital vs. café, ranked as the first factor) and to room acoustic conditions (i.e. STI, ranked as the second factor), language being statistically less relevant (ranked last in Table 6). Furthermore, interaction effects were also important and played a significant role in the differences observed (ranking positions 3-5 out of 6).

Table 6 Ranking of factors showing statistically significant effects over all the attributes tested (from highest effect to lowest effect).

Ranking position	Factor(s)	Number of significant effects
1	Environment	10
2	STI	9
3	STI × Environment	6
4	STI × Language	5
4	STI × Environment × Language	5
5	Environment × Language	3
6	Language	2

Key findings of other semantic attributes

- Speech loudness, speech pleasantness, comfort and environmental pleasantness were significantly correlated with the perceived intelligibility attribute scores of all languages.
- The excitement attribute was the only one showing no significant correlations across all languages.
- Language effects were statistically significant only for the noisiness attribute ($p = 0.047$).
- Soundscape assessment was mostly affected by the environment and room acoustic conditions.
- Interaction effects also played a significant role in soundscape assessment, i.e. independent factors cannot be treated in isolation.

3.2 Principal Component Analysis (PCA)

The interactions between attributes were investigated by using Principal Component Analysis (PCA), in view of extracting meaningful factors (correlated groups of attributes). PCA with a varimax rotation of 11 semantic attributes and the language variable was conducted on data gathered from the 60 participants for each of the 9 cases (3 environments × 3 STI conditions). The components, their attributes, eigenvalues and total variances across the 9 conditions tested are presented in Table 7.

Results show that, for all the conditions tested, the 11 semantic attributes could always be loaded onto 4-5 components. Table 8 actually shows on which components each attribute was loaded for all the conditions tested, in order to identify under which component each attribute tended to be most commonly loaded.

Table 7 PCA results showing the components, their attributes, eigenvalues and total variances across all 9 conditions (3 environments × 3 STIs).

	Component	STI=0.4			STI=0.5			STI=0.6		
		Attributes	Eigenvalue	Variance (%)	Attributes	Eigenvalue	Variance (%)	Attribute	Eigenvalue	Variance (%)
Airport	1	Annoyance, Relaxation, Environment Pleasantness, Comfort, Noisiness	2.91	24.31	Annoyance, Comfort, Noisiness, Relaxation	2.37	19,79	Annoyance, Noisiness, Environment Pleasantness, Relaxation, Comfort	2.93	24.43
	2	Eventfulness, Speech Intelligibility, Language	2.04	17.06	Language, Loudness	1.81	15,14	Excitement, Language	1.49	12.47
	3	Speech loudness, Speech Pleasantness	1.54	12,91	Speech intelligibility	1.62	13,5	Speech Intelligibility, Speech Pleasantness	1.46	12.18
	4	Familiarity, Excitement	1.06	8.83	Eventfulness, Excitement	1.45	12,08	Familiarity	1.42	11.87
Hospital	1	Annoyance, Relaxation, Environment Pleasantness	2.48	20.73	Noisiness, Relaxation, Environment Pleasantness	2.65	22,16	Speech Pleasantness, Annoyance	2.11	17.65
	2	Excitement	1.89	15.75	Speech Intelligibility, Speech loudness	2.33	19,43	Language, relaxation, Eventfulness	1.76	14.74
	3	Speech Loudness	1.6	13.33	Excitement, Language	1.37	11,47	Excitement	1.67	13.98
	4	Familiarity, Speech Intelligibility	1.33	11.1	Familiarity	1.27	10,59	Speech Loudness, Speech Intelligibility	1.47	12.32
	5	Speech Pleasantness	1.26	10.55	N/A	N/A	N/A	Noisiness, Familiarity	1.27	10.59
Café	1	Environment Pleasantness, Comfort, Relaxation, Annoyance	2.68	22.33	Annoyance, Relaxation, Comfort, Environment Pleasantness	3.13	26,15	Excitement, Familiarity	2.19	18.24
	2	Language, Familiarity	1.56	13	Language, Familiarity, Noisiness	2.1	17,54	Relaxation, Environment Pleasantness	1.97	16.48
	3	Noisiness, Speech Pleasantness	1.49	12.43	Eventfulness	1.31	10,93	Language, Speech Intelligibility	1.69	14.1
	4	Eventfulness	1.24	10.37	Speech Intelligibility	1.24	10,33	Eventfulness	1.51	12.59
	5	Speech Intelligibility, Speech Loudness	1.22	10.23	N/A	N/A	N/A	Speech Loudness	1.18	9.87

Overall, these results indicate that:

- Component 1 explains 18-26% of the total variance, and its most common attributes are annoyance, relaxation, comfort and environment pleasantness.
- Component 2 explains 12-19% of the total variance, and its most common attribute is language.
- Component 3 explains 11-14% of the total variance, and its most common attributes are speech intelligibility and speech pleasantness.
- Component 4 explains 9-14% of the total variance, and its most common attributes are familiarity and eventfulness.
- Component 5 explains 10-11% of the total variance, and its most common attribute is speech loudness.

Table 8. Summary table of the semantic attributes and components these attributes are loaded in (PCA results).

	Airport			Hospital			Café		
	STI=0.4	STI=0.5	STI=0.6	STI=0.4	STI=0.5	STI=0.6	STI=0.4	STI=0.5	STI=0.6
Speech intelligibility	2	3	3	4	2	4	5	4	3
Speech loudness	3	2		3	2	4	5		5
Speech pleasantness	3		3	5		1	3		
Noisiness	1	1	1		1	5	3	2	
Annoyance	1	1	1	1		1	1	1	
Relaxation	1	1	1	1	1	2	1	1	2
Comfort	1	1	1				1	1	
Environment pleasantness	1		1	1	1		1	1	2
Eventfulness	2	4				2	4	3	4
Excitement	4	4	2	2	3	3			1
Familiarity	4		4	4	4		2	2	1
Language	2	2	2		3	2	2	2	3

This analysis suggests that perception was largely affected by emotional factors (Component 1), language (Component 2) and speech intelligibility and pleasantness (Component 3). Conversely, familiarity, eventfulness and speech loudness tended to be less influential in describing the conditions tested, as these were mainly loaded onto Components 4 and 5. Furthermore, the results show that Component 1 tended to be the component with most attributes, pointing at strong correlations between the emotional attributes tested (annoyance, relaxation, comfort and environment pleasantness). Finally, it can be noted that relaxation and speech intelligibility were the only attributes loaded onto a component for all the conditions tested.

4 DISCUSSION

This section presents possible reasons for the key findings obtained from the semantic differential analysis and the principal component analysis.

First of all, results pointed out discrepancies between objective and perceived speech intelligibility. The perceived speech intelligibility of English was lower than the objective intelligibility under poor room acoustic conditions (STI = 0.4), and it was lower for Polish under good room acoustic conditions (STI = 0.6). Overall, perceived speech intelligibility was also lower for Mandarin, but higher for Arabic. Noticeably, in the present study, English participants were most sensitive to room acoustic conditions, whilst Arab participants were least sensitive to those.

Methodological differences between the two phases of the research might partly explain English results: white noise was used in the first phase of the research, whilst the background noise samples used in the second phase were representative of a real multi-lingual environment, containing specific distractive noise sources (i.e. public announcement in the airport and phone ringing in the hospital) that are particularly noticeable at STI = 0.4 (i.e. when they are louder). Results suggest that these might have been perceived as more disturbing to English participants, especially considering that noisiness and annoyance showed the strongest correlation with perceived speech intelligibility of English (Table 5). Additionally, the airport contained public announcements that are in English, which might have been more distracting for native English speakers; it is however not clear whether information content played a role, as a poor rating of perceived speech intelligibility of English was also found in the hospital, which did not contain public announcements. Furthermore, it should be noted that the results obtained at the café were more similar to the results of the first phase of the study [2], showing that English was the most intelligible language at STI = 0.4. Arguably, this might be because of the steady background noise sample used in that environment (more comparable to the white noise used in phase 1) and the absence of distractive noise sources. In summary, these results suggest that English participants might be more sensitive to meaningful and distractive sound events that could be caused by a cultural reaction, justifying why the scores at STI = 0.4 at the airport and the hospital were lower than at the café. These results are in line with the research of Jeon et al. [9], who found that ratings of human sounds were significantly different across soundscape assessments made by different cultural groups.

Of particular interest are also the results obtained for Arabic, which showed that Arab participants were the least sensitive to room acoustic conditions when rating speech intelligibility. This would appear to again be due to cultural differences, as Arab participants rated noisiness lower across all environments (Figure 6) as well as annoyance across most environments (Figure 7). This suggests that Arab participants are more tolerant to noise and tend to be less annoyed, therefore being more consistent in their rating of speech intelligibility across different levels of background

noise. This is further demonstrated by the fact that Arabic was the only language that showed no correlation between noisiness and perceived speech intelligibility, as well as no correlation between annoyance and perceived speech intelligibility (Table 5). Small variations in perceived speech intelligibility also explain why Arabic had the lowest number of semantic attributes correlating with perceived speech intelligibility (4 out of 10).

Results showed that language affects perceived intelligibility, although the statistical significance was found to be marginal ($p = 0.051$), with 4 out of 9 cases showing statistically significant differences amongst languages ($p < 0.05$). This is actually higher than what was found for objective sentence intelligibility [2], where only 1 case out of 4 ($STI = 0.4$) showed statistically significant differences between languages. This suggests that perceived speech intelligibility might show larger variations across languages than objective speech intelligibility of sentences. Again, this might be related to cultural factors affecting the rating of participants. However, it should be noted that different STIs were tested in the two phases ($STI = 0.2, 0.4, 0.6$ and 0.8 for phase 1, and $STI = 0.4, 0.5$ and 0.6 for phase 2), so these results are not strictly comparable. Language was also found to significantly affect noisiness ratings ($p = 0.047$), suggesting that noisiness was the best indicator of cultural variations amongst the attributes tested. This is in line with the previous research findings of Yang and Kang [7], as well as a number of other studies reviewed by Marquis-Favre et al. [8].

Although discrepancies were observed between the objective and subjective intelligibility comparisons of different languages, this does not mean that objective vs. perceived intelligibility correlations are not good within each language. In fact, correlations examined within each language were found to be good, with an average correlation coefficient $r = 0.86$, which is comparable to what was found by Cox et al. [3] for participants with normal hearing ($r = 0.82$).

Regarding the soundscape assessment of semantic attributes other than perceived speech intelligibility, results showed that speech loudness, speech pleasantness, comfort and environmental pleasantness were significantly correlated with the perceived intelligibility attribute scores of all languages. Whilst speech loudness is expected to correlate with speech intelligibility (a higher signal-to-noise ratio meaning a higher speech intelligibility), the other attributes represent non-acoustical factors normally ignored in speech intelligibility assessments. This is particularly interesting, as it suggests that pleasantness of the speech and environment, as well as comfort, can all help in improving speech intelligibility. In other words, speech intelligibility based solely on room acoustic parameters might mislead designers, as non-acoustical factors also play a role and should be taken into account. Furthermore, it should be noted that most of the semantic attributes tested correlated with perceived speech intelligibility (English, Polish and Mandarin: 8/10; Arabic: 4/10; see Table 5), highlighting that a wide range of factors affect speech intelligibility. Noticeable exceptions were familiarity and in particular eventfulness, which were found not to correlate with perceived speech intelligibility (with the exception of Mandarin for familiarity).

Semantic analysis also indicated that the soundscape assessment was mostly affected by the environment and room acoustic conditions, rather than the language (Table 6). This is somehow not surprising, as the environments selected were substantially different, and the room acoustic conditions also covered a reasonable range of STI values ($0.4 - 0.6$). Furthermore, statistical analysis highlighted the importance of interaction effects, which point at the interrelation of independent factors (environment vs. STI vs. language) and complexity of the soundscapes examined. In other words, independent factors cannot be treated in isolation.

Principal Component Analysis suggested that perception was largely affected by emotional factors (Component 1), language (Component 2) and speech intelligibility and pleasantness (Component 3). Annoyance, relaxation, comfort and environment pleasantness were the attributes most commonly loaded onto Component 1, which is why this component was labelled as 'emotional factors'. This is a very relevant finding, as it points out that emotional factors were the driving forces in soundscape assessment, once again highlighting the need for going beyond physical and objective factors when designing spaces. This is in line with previous findings of Galbrun and Calarco [20], who identified emotional factors as driving factors in the assessment of soundscapes containing water features.

5 CONCLUSIONS

This study complements previous work that looked at the speech intelligibility of English, Polish, Arabic and Mandarin under different room acoustic conditions [2]. More specifically, the work presented here analysed the perceived speech intelligibility of English, Polish, Arabic, and Mandarin, as well as the soundscape assessment of multilingual environments. Semantic differential analysis of 11 semantic attributes (intelligibility, speech level, speech pleasantness, noisiness, annoyance, relaxation, comfort, environmental pleasantness, eventfulness, excitement, and familiarity) and principal component analysis were conducted for 9 cases representative of a range of environments (airport vs. hospital vs. café) and acoustic conditions ($STI = 0.4, 0.5$ and 0.6). Referring to the objectives listed in section 1, the main findings of the work are:

- 1) Inter-language comparisons based on objective and perceived speech intelligibility can provide different findings. This occurred partly because of methodological differences between the objective and perceptual

tests used, but also because of cultural differences that appeared to have played a significant role in perceived intelligibility ratings. Noticeably, results indicated that:

- English participants were most sensitive to changes in room acoustic conditions, and were most negatively affected by the information content and distracting sounds present in the background noise.
 - Arab participants were least sensitive to changes in room acoustic conditions, and were more tolerant to noise.
- 2) Perceived speech intelligibility was significantly correlated with non-acoustical factor such as speech pleasantness, comfort and environment pleasantness. Furthermore, the 'emotional factors' annoyance, relaxation, comfort and environment pleasantness explained a large portion of the variance in soundscape assessment.
 - 3) Language affected perceived speech intelligibility marginally ($p = 0.051$) and noisiness significantly ($p = 0.047$). Results indicated that:
 - Perceived speech intelligibility showed larger variations across languages than objective speech intelligibility of sentences.
 - Noisiness was the best indicator of cultural variations amongst the attributes tested.

Overall, the study demonstrates that designing for speech intelligibility cannot be solely based on room acoustic parameters, especially in the case of multi-lingual environments. The type of environment, room acoustic conditions, type of background noise, language considered, as well as non-acoustical factors such as comfort and pleasantness can all affect speech communication, and should therefore be considered at the design stage.

Acknowledgement

The authors would like to thank Professor Bernadette O'Rourke and Professor Graham H. Turner for the valuable help provided in the establishment and development of the research.

References

- [1] Martin-Jones M, Blackledge A, Creese A. The Routledge handbook of multilingualism. Oxon, UK: Routledge; 2012.
- [2] Galbrun L, Kitapci K. Speech intelligibility of English, Polish, Arabic and Mandarin under different room acoustic conditions. *Applied Acoustics* 2016; 114: 79-91.
- [3] Cox RB, Alexander GC, Rivera IM. Comparison of objective and subjective measures of speech intelligibility in elderly hearing-impaired listeners. *Journal of Speech and Hearing Research* 1991; 34: 904-915.
- [4] Extra G, Yağmur K. Urban multilingualism in Europe: Mapping linguistic diversity in multicultural cities. *Journal of Pragmatics* 2011; 43(5): 1173-1184.
- [5] Wodak R, Kryzyzanowski M, Forchtner B. The interplay of language ideologies and contextual cues in multilingual interactions: Language choice and code-switching in European Union institutions. *Language in Society* 2012; 41(2): 157-186.
- [6] Schafer RM. The Soundscape: Our Sonic Environment and the Tuning of the World. Rochester, USA: Destiny Books; 1994.
- [7] Yang W, Kang J. Acoustic comfort evaluation in urban open public spaces. *Applied Acoustics* 2005; 66: 211-229.
- [8] Marquis-Favres C, Premat E, Aubrée D. Noise and its effects – A review on qualitative aspects of sound. Part II: Noise and Annoyance. *Acta Acustica United with Acustica* 2005; 91: 626-642.
- [9] Jeon JY, Hong JY, Lavandier C, Lafon J, Axelsson Ö, Hurtig M. A cross-national comparison in assessment of urban soundscapes in France, Korea, and Sweden through laboratory experiments. *Applied Acoustics* 2018; 133: 107-117.
- [10] Peutz VMA. Articulation loss of consonants as a criterion for speech transmission in a room. *Journal of the Audio Engineering Society* 1971; 19(11): 915-919.
- [11] Maddieson I. Consonant-vowel ratio. In: Dryer MS, Haspelmath M, Martin, editors. *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology; 2008. Available online at: <<http://wals.info/chapter/3>> [accessed on 01/02/2019].
- [12] I. Maddieson: Tone. In: Dryer MS, Haspelmath M, Martin, editors. *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology; 2013. Available online at <<http://wals.info/chapter/13>> [Accessed on 16/06/2015].
- [13] EN ISO 9921: Ergonomics – Assessment of speech communication. 2003.
- [14] Jones D. An English Pronouncing Dictionary. Dent, London; 1917.
- [15] Boudraa M, Boudraa B. Twenty lists of ten Arabic sentences for assessment. *Acta Acustica United with Acustica* 2000; 86: 870-882.

- [16] Houtgast T, Steeneken HJM. The modulation transfer function in room acoustics as a predictor of speech intelligibility. *Acustica* 1973; 28: 66-73.
- [17] Long, M. *Architectural Acoustics*, London, UK: Academic Press; 2006.
- [18] Osgood CE, Suci GJ, Tannenbaum PH. *The measurement of meaning*. Champaign, USA: University of Illinois Press; 1957.
- [19] Kang J, Zhang M. Semantic differential analysis of the soundscape in urban open public spaces. *Building and Environment* 2010; 45: 150-157.
- [20] Galbrun L, Calarco FMA. Audio-visual interaction and perceptual assessment of water features used over road traffic noise. *Journal of the Acoustical Society of America* 2014; 136(5): 2609-2620.
- [21] Field A. *Discovering statistics using SPSS*. London, UK: Sage; 2009.
- [22] Shrout P, Fleiss J. Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin* 1979; 86(2): 420-428.
- [23] Kitapci K. *Speech intelligibility in multilingual spaces*. Ph.D. thesis, Heriot-Watt University, Edinburgh, UK; 2016.